UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/660,780	09/12/2003	Nambi Seshadri	58268.00224	5880

32294 7590 05/05/2008

SQUIRE, SANDERS & DEMPSEY L.L.P.
8000 TOWERS CRESCENT DRIVE
14TH FLOOR
VIENNA, VA 22182-6212

EXAMINER	
LERNER, MARTIN	

ART UNIT	PAPER NUMBER
2626	

MAIL DATE	DELIVERY MODE
05/05/2008	PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Office Action Summary	Application No.	Applicant(s)	
	10/660,780	SESHADRI, NAMBI	
	Examiner	Art Unit	
	MARTIN LERNER	2626	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE *3* MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *13 March 2008*.

2a)☒ This action is **FINAL**. 2b)☐ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1 to 21* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1 to 15* is/are rejected.

7)☒ Claim(s) *16 to 21* is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☒ The drawing(s) filed on *27 July 2007* is/are: a)☒ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☐ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☐ Information Disclosure Statement(s) (PTO/SB/08) Paper No(s)/Mail Date _____.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____ .
5) ☐ Notice of Informal Patent Application
6) ☐ Other: _____.

DETAILED ACTION

Claim Rejections - 35 USC § 103

1. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

2. Claims 1 to 3, 5 to 7, 9 to 11, and 13 to 15 are rejected under 35 U.S.C. 103(a) as being unpatentable over *Morris* in view of *Thambiratnam et al. ("Speech Recognition in Adverse Environments using Lip Information")*.

Concerning independent claims 1, 5, and 9, *Morris* discloses a speech recognition method, device, and system, comprising:

"an audio signal receiver configured to receive audio signals from a speech source" – a user speaks to system 100, and system 100 captures the user's speech with speech input unit 104 (column 4, lines 15 to 19: Figures 1 and 2: Block 202); speech is an audio signal;

"a video signal receiver configured to receive video signals from the speech source" – a user speaks to system 100, and system 100 captures the user's image with video input unit 102 (column 4, lines 15 to 19: Figures 1 and 2: Block 202);

"a processing unit configured to process the audio signals and the video signals" – system 100 combines any captured speech or video and proceeds to process the

combined data stream in multi-sensor fusion/recognition unit 106 (column 4, lines 20 to 24: Figures 1 and 2: Block 204);

"a conversion unit configured to convert at least one of the audio signals and the video signals to recognizable information" – system 100 interprets any verbal input using the speech recognition functions of multi-sensor fusion/recognition unit 106; speech recognition is supplemented by visual information captured by video input unit 102, such as any interpreted facial expressions (e.g., lip-reading); a list of spoken words is generated from the verbal input (column 4, lines 25 to 31: Figures 1 and 2: Block 206); spoken words are recognizable information;

"an implementation unit configured to implement a task based on the recognizable information" – system 100 provides a response based upon whether the user has asked a question or made a statement; if a user has asked a question, then system 100 searches knowledge database 116 for a response to the objective question; a user may ask: "What is the weather in Phoenix, today?"; system 100 retrieves an answer, and the information is communicated as output via computer monitor and speakers (column 4, line 56 to column 5, line 24: Figure 3: Blocks 306, 308, 310, 312, 322); responding to a question by searching a knowledge database for a weather report in Phoenix, and outputting the weather report, is equivalent to implementing a task.

Concerning independent claims 1, 5, and 9, the only elements arguably omitted by *Morris* are "detecting if the audio signal can be processed", processing the audio signals "if it is detected that the audio signals can be processed", and processing the video signals "if it is detected that at least a portion of the audio signal cannot be

processed". *Morris* discloses processing both the audio and video signals for multi-sensor fusion, so that better recognition can be obtained from speech input and video input. Fundamentally, one having ordinary skill in the art would readily understand that a speech recognizer that utilizes both audio and video for purposes of recognition would utilize the video if the quality of the audio information is poor, and utilize the audio if the quality of the audio information is good.

Concerning independent claims 1, 5, and 9, specifically, *Thambiratnam et al.* teaches speech recognition in adverse environments, where asynchronous integration merges the results of two systems together to produce a combined probability:

$$P_c = \lambda P_A + (1 - \lambda P_V),$$

where P_A represent the acoustic score from the acoustic subsystem, P_V represents the visual scores from the video subsystem, and λ is a weighting parameter that depends on the signal-to-noise (SNR) ratio. (§4.2 Asynchronous Integration: Pages 150 to 151: Figure 3) Moreover, Figure 4 illustrates performance accuracy as a function of SNR, where the visual subsystem performs at the same error rate of approximately 85%, regardless of the SNR, but that the acoustic subsystem performance degrades rapidly as the SNR decreases. In fact, Figure 4 shows that for SNR < 5, a video subsystem will provide better accuracy than any of the acoustic subsystems of Mel-Cepstral, RASTA, or Mel-RASTA. (§5.1 Individual Sub-System Performance: Page 151: Figure 4) (Setting the weighting parameter, λ=0, corresponds to processing only the video signals.) Thus, one skilled in the art would have found it "obvious to try" processing the video signals based on a detection that at least a portion of the audio signal cannot be

processed due to a low signal-to-noise ratio as taught by *Thambiratnam et al.* It would have been obvious to one having ordinary skill in the art to process the video signals based on a detection that at least a portion of the audio signal cannot be processed as suggested by *Thambiratnam et al.* in a multi-sensor fusion/recognition unit of *Morris* for a purpose of improving an accuracy of speech recognition in adverse environments for conditions of low signal-to-noise ratios.

Concerning independent claims 13 to 15, similar considerations apply as to independent claims 1, 5, and 9. Implicitly, the signal-to-noise ratio must be a function of time, and the audio and video segments coincide in time, so that *Thambiratnam et al.* would process the audio and video as segments coinciding in time.

Concerning claims 2, 6, and 10, *Morris* discloses that video input unit 102 receives face/voice expressions and interpreted facial expressions including lip-reading (column 4, lines 27 to 30: Figures 1 and 2).

Concerning claims 3, 7, and 11, *Morris* discloses that, in one embodiment, processing by multi-sensor fusion recognition unit 106 is split into three parallel processes to minimize time of processing (column 4, lines 20 to 24: Figures 1 and 2).

3. Claims 4, 8, and 12 are rejected under 35 U.S.C. 103(a) as being unpatentable over *Morris* in view of *Thambiratnam et al. ("Speech Recognition in Adverse*

Environments using Lip Information") as applied to claims 1, 5, and 9 above, and further in view of *Bakis et al*.

Morris does not expressly disclose a storage unit for storing the audio signals and the video signals to a destination source, and a transmitter for sending the audio signals and the video signals to a destination source. However, it is well known to operate biometric identification via a client/server network, where biometric data is stored on a server, and biometric data is collected locally but compared to stored biometric data on the server. *Bakis et al*. teaches an analogous art method and apparatus for recognizing the identity of individuals by a speaker recognition system and a lip classifier, where biometric attributes are pre-stored for later retrieval so that they may be compared. Further, a server is included for interfacing with a plurality of biometric recognition systems to receive requests for biometric attributes therefrom and transmit biometric attributes thereto. The server has a memory device for storing the biometric attributes. (Column 8, Line 47 to Column 9, Line 16) Objectives are to provide a significant increase in the degree of accuracy of recognition and to provide a significant reduction in fraudulent or errant access to a service and/or facility. It would have been obvious to one having ordinary skill in the art to store and send biometric attributes to a server ("a destination source") as taught by *Bakis et al*. in a method, device, and system for combining audio and video signals of *Morris* for purposes of increasing accuracy of recognition and reducing fraudulent access.

Allowable Subject Matter

4. Claims 16 to 21 are objected to as being dependent upon a rejected base claim, but would be allowable if rewritten in independent form including all of the limitations of the base claim and any intervening claims.

5. The following is a statement of reasons for the indication of allowable subject matter:

Regarding claims 16 to 18, the prior art of record does not disclose or reasonably suggest the limitation of defining an error threshold, comparing a number of detected errors in an audio signal with the threshold, and determining that the audio signals cannot be processed if the number of errors equals or exceeds the threshold. The prior art of record suggests a parameter involving a signal-to-noise ratio for an audio signal to determine that the quality of the audio signal is sufficient to obtain good accuracy for speech recognition, but does not compare a number of errors with a threshold.

Regarding claims 19 to 21, the prior art of record does not disclose or reasonably suggest the limitation of indicating to the user if the video image is not detected. *Brunelli et al*. teaches an attention module that waits until a video image has stabilized before prompting a user to speak. However, prompting a user to speak when a video image is detected is not the same as indicating to the user that a video image is not detected. Specifically, a user may be alerted to change his/her position in a video field if a positive indication is given to the user that the video image is not detected.

Response to Arguments

6. Applicant's arguments filed 13 March 2008 have been fully considered but they are not persuasive.

Applicant argues that a combination of *Morris* and *Thambiratnam et al.* fails to disclose or suggest all of the features of independent claims 1, 5, 9, and 13 to 15. Moreover, Applicant submits that it would not have been "obvious to try" processing the video signals based on a detection that at least a portion of the audio signal cannot be processed. Applicant says that neither *Morris* nor *Thambiratnam et al.* provides a teaching of "processing the video signals if it is detected that at least a portion of the audio signal cannot be processed". Applicant notes that *Thambiratnam et al.* discloses integration of video and audio is performed by one of two primary methods of integration, which are direct integration and asynchronous integration. For direct integration, the audio and video vectors are combined as input to a recognizer, and for asynchronous integration, the data is merged based on the two results which are calculated independent of one another, so that the results can be independently determined. However, Applicant states that, in the direct integration and asynchronous integration of *Thambiratnam et al.*, there is still no disclosure of "processing the video signals if it detected that at least a portion of the audio signal cannot be processed". Furthermore, Applicant maintains that the case law standard of what is "obvious to try" is misapplied in the Office Action, and is not relevant to the subject matter of the claims. These arguments are traversed.

Firstly, Applicant disregards the technical discussions about how the signal-to-noise ratio (SNR) affects recognizer performance of the acoustic subsystem as taught by *Thambiratnam et al.* On Page, 151, Left Column, §5.1 Individual Sub-system performance, *Thambiratnam et al.* states, "The visual subsystem obviously performs at the same error rate regardless of the SNR. The results are found in Figure 4. The acoustic sub-system performance degrades rapidly as the SNR decreases." Measuring the signal-to-noise ratio (SNR), and finding a low signal-to-noise ratio, is equivalent to "detecting if the audio signal can be processed" and "if it is detected that at least a portion of the audio signal cannot be processed". A low signal-to-noise ratio (SNR) implies that the acoustic sub-system cannot process the audio signal. On Page 151, Left Column, §4.2 Asynchronous Integration, moreover, *Thambiratnam et al.* clearly states, "λ is a weighting parameter that depends on the SNR." Given a low signal-to-ratio (SNR), then, $\lambda \rightarrow 0$, $\lambda P_A \rightarrow 0$, and $(1 - \lambda)P_V \rightarrow P_V$, so that a combined probability, $P_c = P_V$. Equivalently, this just says that speech recognition is performed with only the video signal if the signal-to-noise ratio (SNR) is low, when the audio signal cannot be processed.

The basic standard to support an obviousness rejection under 35 U.S.C. §103(a) is whether the claims would have been obvious as a whole to one having ordinary skill in the art. Here, those having ordinary skill in the art of speech recognition would know that the whole point of supplementing acoustic speech recognition with video speech recognition is to improve the results and increase the accuracy over utilizing either one of only audio or only video alone to perform speech recognition. Thus, it is

presupposed that if either one of the audio signal or the video signal is suboptimal, then the other of the two will supplement the results and provide the best possible speech recognition under the circumstances. It is maintained that this rationale is based not on hindsight, but on the fundamental reason for employing both the audio signal and the video signal for speech recognition in the first place. Granted, there is no clear teaching in the prior art of detecting if the audio signals can be processed, and processing the video signals based on a detection that at least a portion of the audio signal cannot be processed in a manner given in a form of a flow chart by Applicant's Figure 2. Still, some things are so obvious that they are not written down as such. One having ordinary skill in the art would appreciate that, given considerations of the effects of the signal-to-noise ratio (SNR) on the audio signal as set forth by *Thambiratnam et al.*, and the underlying motivation for utilizing a video signal to perform lip reading to supplement a conventional method of utilizing an audio signal to perform speech recognition, it would follow to process the video signals based on a detection that at least a portion of the audio signal cannot be processed.

Secondly, Applicant's analysis of the standard for "obvious to try" is met by a combination of *Morris* and *Thambiratnam et al.* The first (1) requirement is met because one of ordinary skill in the art would have recognized the need for substituting video for audio if the audio cannot be processed. Indeed, the entire reason for utilizing both audio and video for speech recognition, instead of conventionally just utilizing the audio for speech recognition, is that the video provides additional information that improves recognition performance. The second (2) requirement is met because there are a finite

number of identified, predictable solutions to the problem. Clearly, one can utilize only the audio for speech recognition, only the video for speech recognition, or some combination of the audio and the video for speech recognition. The possibilities for utilizing the audio and the video signals are limited to a small set of combinations. Given the first (1st) example, provided by Applicants, of a pharmaceutical drug that was held to have a known result that was predictable based on a list of fifty-three possibilities, there are certainly a smaller set of possible solutions here to the problem of combining audio signals and video signals for speech recognition than a list of fifty-three. Similarly, given the third (3rd) example, provided by Applicants, that there are only a few known ways to isolate a nucleic acid molecule, which were found to be "obvious to try", here there are essentially only a few combinations of ways to utilize audio signals and video signals in speech recognition: one can only use the audio signals, one can only use the video signals, or one can use some combination of the audio and video signals.

Applicant has presented arguments traversing the rejection of claims 19 to 21 under 35 U.S.C. §103(a) as being obvious over *Morris* in view of *Thambiratnam et al.*, and further in view of *Brunelli et al*. These arguments are persuasive, and the rejection is withdrawn.

Therefore, the rejections of claims 1 to 3, 5 to 7, 9 to 11, and 13 to 15 under 35 U.S.C. §103(a) as being unpatentable over *Morris* in view of *Thambiratnam et al.*, and of claims 4, 8, and 12 under 35 U.S.C. §103(a) as being unpatentable over *Morris* in view of *Thambiratnam et al.*, and further in view of *Bakis et al*., are proper.

Conclusion

7. **THIS ACTION IS MADE FINAL.** Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the mailing date of this final action.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Martin Lerner whose telephone number is (571) 272-7608. The examiner can normally be reached on 8:30 AM to 6:00 PM Monday to Thursday.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, David R. Hudspeth can be reached on (571) 272-7843. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for

published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/Martin Lerner/
Primary Examiner, Art Unit 2626
May 1, 2008